

PSYCHOMETRIC PROPERTIES OF THREE ADDITION TASKS
WITH DIFFERENT RESPONSE REQUIREMENTS

Louis E. Banderet, PhD, Barbara L. Shukitt, B.A.,
SSG Michael A. Walthers

U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760-5007 USA

Robert S. Kennedy, PhD
Essex Corporation
1040 Woodcock Road, Orlando, FL 32803 USA

Alvah C. Bittner Jr., PhD
Analytics, Inc.
2500 Maryland Rd., Willow Grove, PA 19090 USA

Gary G. Kay, PhD
Georgetown University Hospital
3800 Reservoir Road NW, Washington DC 20007-2197 USA

DTIC
ELECTE
31 MAR 1989
S E D

AD-A206 939

PAPER PRESENTED AT AND PUBLISHED IN PROCEEDINGS 30TH ANNUAL
MEETING MILITARY TESTING ASSOC., ARLINGTON, VA, NOV. 1988.

ABSTRACT

The psychometric properties of performance tasks are influenced by many variables, e.g., task demands, sensory modality tested, requirements for transfer of skills or new learning, and the subject's output response. Previously, we administered two Addition Tasks, where each problem required mentally summing three two-digit numbers. The format, spatial layout, and problem characteristics were similar; however, each task required a different output response (writing each answer on paper versus entering each with number keys on the top row of the computer keyboard). Although we gave subjects "touch typing" practice on the number keys, subjects responded 65% as fast on the automated task and its differential stabilities were less. To overcome the limitations of the computerized Task, we developed another Addition Task for computer administration that required only a dichotomous response ("Right" or "Wrong"). Problem format was similar to earlier versions, except each problem had a suggested sum and it was wrong on 50% of the problems. Subjects mentally calculated each sum, determined if each suggested sum was right or wrong, and entered their answer on the keyboard.

Recently, we evaluated the three tasks in a study which manipulated environmental stressors. The first computer task, requiring entry of the sum with the number keys on the computer, was psychometrically inferior to the other two. The second computer task was superior to even the paper and pencil task although its problem solving rate was 20% lower. These results demonstrate how relatively subtle differences in the response requirements of a task may dramatically influence its psychometric properties. Such findings suggest that traditional measures of neurological, cognitive, or perceptual status must be implemented cautiously on alternate media.

This document has been approved
for public release and sale its
distribution is unlimited.

89 3 30 054

Accession For	NTIS GRA&I	<input checked="" type="checkbox"/>
	DTIC TAB	<input type="checkbox"/>
	Unannounced	<input type="checkbox"/>
	Justification	<i>from 50 per</i>
By		
Distribution/		
Availability Codes		
Avail and/or Special		20
Dist		PH

QUALITY
INSPECTED
4

Technical advances in personal computers make them very useful for automated testing of questionnaires, personality inventories, and performance tasks. When traditional assessment instruments are adapted for administration by computer, changes in task characteristics or output responses are sometimes inevitable. For example, writing numbers on a paper and pencil task versus entering them with number keys on a task administered by computer. Hence, implementing tasks on the computer may change their psychometric properties. Testing tradition implies that the computerized instrument should be validated. Indeed, recent evaluations with systematic criteria suggest such caution is warranted (Bittner, Carter, Kennedy, Harbeson, & Krause, 1986; Bittner, Smith, Kennedy, Staley, & Harbeson, 1985; Farrell, 1983, and Smith, Krause, Kennedy, Bittner, & Harbeson, 1983).

Recently, we adapted a paper and pencil Addition Task for administration on a portable computer. We found that it seemed insensitive to experimental effects and its response rate was 65% of that for the paper and pencil task. Therefore, we developed another Addition Task to overcome these shortcomings. The purpose of this investigation was to determine the psychometric properties of three Addition Tasks with different response requirements.

METHOD

Subjects---Twenty medical research volunteers from Fort Detrick, MD, and the Natick, MA, Research Development and Engineering Center were subjects. All were given physicals; no subjects had medical histories that would contraindicate altitude and cold exposure. Human subjects participated in these studies after giving their free and informed voluntary consent. Investigators adhered to AR 70-25 and USAMRDC Regulation 70-25 on Use of Volunteers in Research.

Psychometric Instruments---Three Addition Tasks were evaluated (see Figure 1). All tasks were timed and required summing three two-digit numbers to solve each problem; however, the output response differed for the three tasks. The administration medium also differed for some.

The P-SUM Addition Task was a traditional paper and pencil task, generated by computer (Carter & Sbisa, 1982), with 60 problems per page. A subject could change an answer anytime during the testing interval.

The C-SUM and C-RorW Addition Tasks were administered on a GRiD Compass portable computer with a high contrast electro-

luminescent display. The spatial arrangement of the 40 problems per display for these tasks was similar. The C-SUM task required entering each three-digit sum with the number keys on the top row of the keyboard. The cursor's position below a problem prompted for the sum's digits in a specific order. The rightward digit of each sum was entered first, then the leftward, and finally the center digit. An arithmetic or entry error could not be changed after a digit was entered on the keyboard.

The C-RorW Addition Task presented a suggested sum for each problem which was wrong on 50% of the problems. Subjects mentally calculated each problem's sum and decided if each suggested sum differed from the correct sum. The task was programmed for deviations of +1 in the 100, 10, or 1-digit of the sums on 2.5, 45, or 2.5% of all problems, respectively. Subjects pressed the 'F' or 'G' key if the suggested sum was correct; they pressed the 'H' or 'J' key if it was wrong. Like the C-SUM task, a response could not be changed after a key was pressed.

Procedures---This study was part of a larger investigation to determine if an amino acid, tyrosine, prevents some of the adverse behavioral effects induced by environmental stressors. Specifically, on three occasions two groups of 10 subjects were exposed to 4700 m of simulated high altitude and 17°C for 7 h and investigated with a repeated-measures design. Conditions were blinded as to administration of placebo and tyrosine (85 mg/kg and 170 mg/kg). Two other occasions, subjects were given placebo and tested at 500 m and 22°C. Before the experiment, subjects practiced each task 15 times (4 min per administration). Subjects were trained to work quickly with less than an 8% error rate. Feedback was given during training to facilitate rapid acquisition of the tasks and enhance motivation.

During experimental testing, subjects also performed each task for 4 min. Physiological and biochemical measures and an extensive symptom, mood, and performance battery were administered 90-420 min after ascent to simulated high altitude. The Addition Tasks were administered from 320-350 min, immediately after the third blood sample. The tasks were always investigated in the same order: C-RorW, P-SUM, and C-SUM. Map Compass, another task that we developed, and Number Comparison (Carter & Sbisà, 1982) were administered before and after the P-SUM task, respectively, to separate the three Addition Tasks. Sea level performance values were the average of the two sessions at 500 m; 4700 m values were from the day when the placebo was given to each subject.

Performance on each task was specified as number of problems correct per minute. To discourage careless responding, each error was doubled in the calculation of this index. To determine task sensitivity, a z score was calculated for each task, reflecting both the magnitude and variability of measured altitude effects. To obtain another estimate of the relative sensitivities of the tasks, z scores (altitude effects) for each task were calculated for 10 samples randomly resampled from the original subject population. Each sample ($N = 10$) was drawn without replacement; however, the parent population ($N = 20$) was always reconstituted before another sample was selected.

RESULTS

Figure 2 shows scatterplots with correlation coefficients relating each Addition Task with each other task for the sea level + 22°C (500 m) and simulated high altitude + 17°C (4700 m) conditions. High altitude conditions produced significant impairments in performance on all three tasks ($p < .0001$) which were usually evident in the scatterplots as a leftward and downward displacement of data points for the 4700 m condition. All correlations in Figure 2 were statistically significant. The C-RorW task was consistently more highly correlated with the P-SUM task for both environmental conditions than the C-SUM task was. The two computerized versions of the Addition task were more similar to each other for both environmental conditions than either was to the P-SUM task. Performance rates differed on the three tasks; the C-SUM task produced the slowest rates.

Table I shows characteristics of the three Addition Tasks for the 500 and 4700 m conditions. Performance rates were greatest for the P-SUM task. At 500 m, rates on the C-RorW task were 80% as great; those on the C-SUM task were 60%.

P-SUM			
43	75	52	MEDIUM: PAPER & PENCIL
38	26	73	RESPONSE: WRITE SUM
28	54	88	
101	195		
C-SUM			
34	82	50	MEDIUM: COMPUTER
79	82	92	RESPONSE: ENTER SUM
46	97	11	
119			
C-RorW			
99	63	55	MEDIUM: COMPUTER
32	29	25	RESPONSE: ENTER R or W
49	35	26	
180	117	116	
R	W		

Figure 1. Sample problems, administration media, and output responses for three Addition Tasks.

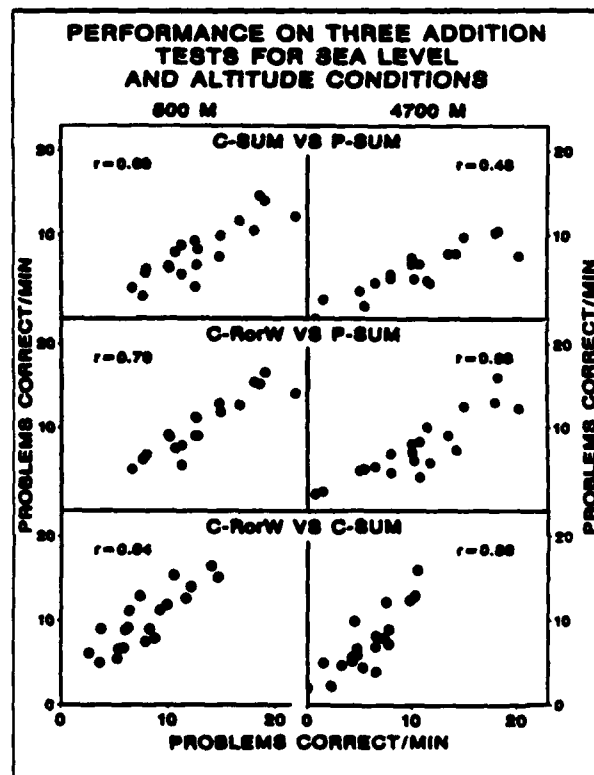


Figure 2. Scatterplots relating performance for each Addition Task with every other task at sea level + 22°C and simulated high altitude + 17°C conditions.

CHARACTERISTICS OF ADDITION TASKS AT 500 AND 4700 M

CRITERION	STATISTIC	P-SUM	C-SUM	C-RorW
500 M				
BASELINE RATES (CORRECT/MIN)	MEAN	12.87	7.97	10.11
	SIGMA	4.23	3.38	3.49
MINUTES PRACTICE (4 MIN/ADMIN)	MEAN	520	540	520
TASK DEFINITION (ADMINS 5 & 6)	PEARSON r	>0.85	>0.72	>0.85
RELIABILITY (4700 VS 500 M)	PEARSON r	0.76	0.70	0.93
4700 M				
ALTITUDE EFFECT (Δ CORRECT/MIN)	MEAN	-2.39	-2.22	-2.63
	SIGMA	1.97	1.85	1.84
	z SCORE	-1.22	-1.20	-1.43

NORMALIZED ALTITUDE EFFECTS					
RANKS			z SCORES		
P-SUM	C-SUM	C-RorW	P-SUM	C-SUM	C-RorW
3	2	1	-1.12	-1.26	-1.64
3	2	1	-0.80	-0.96	-1.15
2	3	1	-1.08	-0.87	-1.31
2.5	2.5	1	-1.50	-1.50	-1.56
2	3	1	-1.24	-0.85	-2.04
1	2	3	-1.75	-1.65	-1.54
3	2	1	-1.42	-1.73	-1.82
3	1	2	-1.17	-1.69	-1.57
2	3	1	-1.29	-1.20	-1.59
1	2	3	-1.55	-1.12	-1.08
2:10	1:10	7:10	PROPORTION WITH LARGEST EFFECT		

Table I. Psychometric characteristics of three Addition Tasks for sea level + 22°C and simulated high altitude + 17°C conditions.

Table II. Sensitivities of three Addition Tasks determined with data resampled from the original subject population.

Variability was also greatest for the P-SUM task; it was less for the two computer administered tasks. The C-SUM task required twice as much practice as the other tasks to achieve comparable task stability ($> .80$). Task reliability for the experimental and control administrations was greatest for the C-RorW task. Task definition (test-retest reliability) was greatest for the P-SUM and C-RorW tasks. Table I also shows the mean altitude effects and standard deviations for each task. The C-RorW task was the most sensitive task as inferred from z score magnitudes; the P-SUM and C-SUM tasks were less sensitive.

Table II shows z scores and their ranks (1 = most sensitive) from our resampling procedure. With 10 samples, P-SUM was most sensitive twice, C-SUM was once, and C-RorW was seven times.

DISCUSSION

Performances on three Addition Tasks with differing response requirements were compared under control and stressful environmental conditions. This study demonstrated that our first computer implementation of a useful paper and pencil task (P-SUM) resulted in an automated task (C-SUM) with lower response rates, reduced sensitivity to experimental effects, less task definition, and greater practice requirements. It is of interest that this Addition Task is currently implemented in varied performance batteries. We suspect that entering numbers on a computer keyboard was awkward and incompatible with rapid responding since few of our subjects could "touch type".

To overcome the limitations of the initial computerized Addition Task, we developed a second Addition Task (C-RorW) which required a dichotomous response. Although the problem solving rate of the second computerized task was less than expected, its sensitivity and other psychometric properties were superior to the paper and pencil task. We suspect our second computerized task was more sensitive because it was more abstract and complex, e.g., subjects determined the sum for each problem, compared it to the suggested sum, and decided which response key to press.

Implementation of an Addition Task with an appropriate subject output response produced a computerized task that was psychometrically superior to its paper and pencil counterpart. In contrast, choice of the traditional output response for our first computerized task yielded the task that was the most inferior. These data illustrate the importance of evaluating tasks when they are changed or adapted to alternative media. They also demonstrate how relatively subtle differences in the response requirements of a task may dramatically influence its psychometric properties. Such findings suggest that traditional measures of neurological, cognitive, or perceptual status must be implemented cautiously on alternative media.

REFERENCES

Bittner, A.C., Jr., Carter, R.C., Kennedy, R.S., Harbeson, M.M., & Krause, M. (1986). Performance evaluation tests for environmental research: Evaluation of 114 measures. Perceptual Motor Skills, 63, 683-708.

Bittner, A.C., Smith, M.G., Kennedy, R.S., Staley, C.F., & Harbeson, M.M. (1985). Automated portable test (APT) system: Overview and prospects. Behavior Research Methods, Instruments, & Computers, 17(2), 217-221.

Carter, R.C., & Sbisá, H. (1982). Human performance tests for repeated measurements: Alternate forms of eight tests by computer (Report No. NBDL8213003). New Orleans, LA: Naval Biodynamics Laboratory.

Farrell, A.D. (1983). When is a computerized assessment system ready for distribution? Some standards for evaluation. Computers in Psychiatry/Psychology, 5(1), 9-11.